DTU Library

Is your data correct? Validating and improving data collected in smart water networks

Kirstein, Jonas Kjeld; Høgh, Klavs; Borup, Morten; Rygaard, Martin

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Is your data correct? Validating and improving data collected in smart water networks

by Jonas Kjeld Kirstein | Klavs Høgh | Morten Borup | Martin Rygaard | Technical University of Denmark | NIRAS A/S | Technical University of Denmark | Technical University of Denmark

Abstract Id: 61
Submitted: December 1, 2017
Event: Nordic Drinking Water Conference 2018
Topic: Distribution systems

Is your data correct? Validating and improving data collected in smart water networks

Jonas Kjeld Kirstein*[,a], Klavs Høgh[b], Morten Borup[a] and Martin Rygaard[a]

[a]Department of Environmental Engineering, Technical University of Denmark, Bygningstorvet B115, 2800 Kgs. Lyngby, Denmark; [b]NIRAS A/S, Sortemosevej 19, 3450 Allerød, Denmark

*Corresponding author; e-mail: jkir@env.dtu.dk; phone: +45 45 25 14 79

Intended for presentation

Within recent years, drinking water utilities have increased the installation of data collecting devices throughout their water distribution networks. The purpose of this data collection is to improve systems operation and to help secure drinking water safety and reliability even further. The growing data-dependency requires maintenance of high data integrity and monitoring of data validity. However, data validity and integrity testing is a lesser-described process step in research and water systems operation. As the collected data can suffer from errors that can severely reduce its value, it is of importance to quality assure the data automatically and in real time to avoid long periods with erroneous data. Most applications and studies within distribution system modelling assume flawless data and conduct analyses based on simple filtering of outliers and blank spaces. But, what if filtered outliers represent true extremes? What if time settings change or drift for some meters in the collected data? Can anomalies easily and reliably be replaced with estimated data? To improve the reliability and usability of the collected time series data, we propose a

method for quality assurance and emulation of raw data in real time. The method combines well-known and novel data analysis methods to produce an operational time series database. The method also constructs a malfunction and indicator database to highlight problems within data acquisition practices. A demonstration of the method was carried out for three utilities in Denmark including 232 meter data sets covering on average 32 months of data. We found that various data errors are frequently occurring and that approximately 10% of measurements collected across the utilities were missing or flawed. This demonstrated that there is 1) a high need for data emulation and 2) that issues within data acquisition of meter data appear independently across the three analysed utilities. However, the number and type of anomalies in the collected data varied greatly between the utilities and the analysed parameters. For example, whereas around 5% of the collected flow and pressure data in two utilities were classified as flatline anomalies (i.e. the exact same values were measured continuously over time), more than 30% of the flow data were categorized as flatline segments in the third utility. In current operation schemes, such errors are easily overlooked because they are masked in applications dependent on hourly or daily aggregates. Our method provides an automated routine for identifying the errors in real time. Moreover, we identified that certain anomalies were correlated throughout the network, which highlighted issues within data collection practice. Identification of systematic anomalies allows operators to suggest and implement improvements to future data collection procedures. Finally, we trained artificial neural networks based on measurements from correlated meters in the network to reconstruct hourly data. The important outcome of our quality assurance and emulation is the construction of an operational database that provides utilities with validated and uniform data streams. The validated data can now be used as trusted input for hydraulic and water quality modelling, for example in leakage management and source tracking. Also, the identification and illustration of correlated errors can be used to reduce the frequency and duration of erroneous measurements over time.